



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2022

Interactions in fixed effects regression models

Giesselmann, Marco ; Schmidt-Catran, Alexander W

Abstract: An interaction in a fixed effects (FE) regression is usually specified by demeaning the product term. However, algebraic transformations reveal that this strategy does not yield a within-unit estimator. Instead, the standard FE interaction estimator reflects unit-level differences of the interacted variables. This property allows interactions of a time-constant variable and a time-varying variable in FE to be estimated but may yield unwanted results if both variables vary within units. In such cases, Monte Carlo experiments confirm that the standard FE estimator of $x \cdot z$ is biased if x is correlated with an unobserved unit-specific moderator of z (or vice versa). A within estimator of an interaction can be obtained by first demeaning each variable and then demeaning their product. This “double-demeaned” estimator is not subject to bias caused by unobserved effect heterogeneity. It is, however, less efficient than standard FE and only works with $T < 2$.

DOI: <https://doi.org/10.1177/0049124120914934>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-187386>

Journal Article

Supplemental Material

Originally published at:

Giesselmann, Marco; Schmidt-Catran, Alexander W (2022). Interactions in fixed effects regression models. *Sociological Methods Research*, 51(3):1100-1127.

DOI: <https://doi.org/10.1177/0049124120914934>

Appendix

A1. Transformation of a demeaned term with one time-varying and one time-invariant factor

Let z be constant within units and, therefore, $z_{it} = \bar{z}_i$ for all $\{i, t\}$. In this case, the demeaned interaction term (3) can be written as

$$(5) \quad (\bar{z}_i x_{it}) - \frac{\sum_{t=1}^{T_i} \bar{z}_i x_{it}}{T_i}.$$

For each measurement $\{i, t\}$, \bar{z}_i is a factor in all added terms in the subtrahend of equation (5). Therefore, it can be factored out, and we obtain

$$(5.1) \quad (\bar{z}_i x_{it}) - (\bar{z}_i \sum_{t=1}^{T_i} \frac{x_{it}}{T_i}).$$

Factoring again \bar{z}_i yields

$$(5.2) \quad \bar{z}_i (x_{it} - \sum_{t=1}^{T_i} \frac{x_{it}}{T_i}).$$

As the sum of all unit-specific measures divided by the number of unit-specific measures establishes the unit-specific mean, (5.2) can be written as

$$(5.3) \quad \bar{z}_i (x_{it} - \bar{x}_i)$$

or in short as

$$(6) \quad \bar{z}_i \ddot{x}_{it}.$$

A2. Transformation of a demeaned term with two time-varying factors

We rewrite the demeaned interaction term (3) as

$$(7) \quad (\bar{z}_i + \ddot{z}_{it})(\bar{x}_i + \ddot{x}_{it}) - \frac{\sum_{t=1}^{T_i} (\bar{z}_i + \ddot{z}_{it})(\bar{x}_i + \ddot{x}_{it})}{T_i}.$$

By expanding the product in (7), we obtain

$$(7.1) \quad \bar{z}_i \bar{x}_i + \bar{z}_i \ddot{x}_{it} + \bar{x}_i \ddot{z}_{it} + \ddot{z}_{it} \ddot{x}_{it} - \frac{\sum_{t=1}^{T_i} \bar{z}_i \bar{x}_i + \bar{z}_i \ddot{x}_{it} + \bar{x}_i \ddot{z}_{it} + \ddot{z}_{it} \ddot{x}_{it}}{T_i}.$$

By fractional arithmetic, (7.1) can be transformed into

$$(7.2) \quad \bar{z}_i \bar{x}_i + \bar{z}_i \ddot{x}_{it} + \bar{x}_i \ddot{z}_{it} + \ddot{z}_{it} \ddot{x}_{it} - \frac{\sum_{t=1}^{T_i} \bar{z}_i \bar{x}_i}{T_i} - \frac{\sum_{t=1}^{T_i} \bar{z}_i \ddot{x}_{it}}{T_i} - \frac{\sum_{t=1}^{T_i} \bar{x}_i \ddot{z}_{it}}{T_i} - \frac{\sum_{t=1}^{T_i} \ddot{z}_{it} \ddot{x}_{it}}{T_i}.$$

The summations in the numerators contain common factors: unit means. Factoring these, we obtain

$$(7.3) \quad \bar{z}_i \bar{x}_i + \bar{z}_i \bar{x}_{it} + \bar{x}_i \bar{z}_{it} + \bar{z}_{it} \bar{x}_{it} - \bar{z}_i \bar{x}_i - \bar{z}_i \frac{\sum_{t=1}^{T_i} \dot{x}_{it}}{T_i} - \bar{x}_i \frac{\sum_{t=1}^{T_i} \dot{z}_{it}}{T_i} - \frac{\sum_{t=1}^{T_i} \dot{z}_{it} \dot{x}_{it}}{T_i}.$$

Because the sum of demeaned values on the unit level is zero, equation (7.3) can be reduced to

$$(8) \quad \bar{z}_i \dot{x}_{it} + \bar{x}_i \dot{z}_{it} + \dot{z}_{it} \dot{x}_{it} - \frac{\sum_{t=1}^{T_i} \dot{z}_{it} \dot{x}_{it}}{T_i}.$$

A3. Fixed Effects Error Terms in Presence of Effect Heterogeneity

Unit-specific effect heterogeneity remains in the error term of a fixed effects regression. This is empirically revealed by the option to specify this heterogeneity systematically (through interactions of unit-specific with time-dependent variables), but can also be shown algebraically:

May the following equation describe the data generating process:

$$(A3.1) \quad y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 z_{it} + \beta_3 (z_{it} x_{it}) + w, \quad w = u_{0i} + u_{1i} x_{it} + u_{2i} z_{it} + e_{it}$$

where u_{1i} is unobserved effect heterogeneity of x between units. For reasons of simplicity, we assume that there is no unobserved effect heterogeneity of z .

An FE transformation of (A3.1) yields the error-term

$$(A3.2) \quad w_{fe} = (u_{0i} - \bar{u}_{0i}) + (u_{1i} x_{it} - \bar{u}_{1i} \bar{x}_{it_i}) + (e_{it} - \bar{e}_i),$$

with $\bar{u}_{1i} \bar{x}_{it_i}$ representing the unit-specific mean value of the product of the unit-specific moderator with the observed variable x . As u_{1i} are constant within units, for each it it is

$$(A3.3) \quad \bar{u}_{1i} \bar{x}_{it_i} = \frac{\sum_{t=1}^T u_{1i} x_{it}}{Tn} = \frac{u_{1i} \sum_{t=1}^T x_{it}}{Tn} = u_{1i} \frac{\sum_{t=1}^T x_{it}}{Tn} = u_{1i} \bar{x}_i,$$

and therefore

$$(A3.4) \quad w_{fe} = u_{1i} (x_{it} - \bar{x}_{it}) + e_{it}.$$

The transformed error-term reveals that unit specific heterogeneity in the effect of x remains in the error of a (standard) fixed effects regression.

A4. Implementation of double demeaning via Stata & Hausman test

```
*****
* This code produces a within-interaction estimator of two time-varying variables *
* x and z in a fixed effects framework ("double demeaned interaction estimator"). *
* It compares the coefficient with the estimator obtained from a conventional FE- *
* approach and computes a hausman-test to detect a possible bias in this      *
* conventional estimator                                                    *
*****
* To run smoothly, this ado needs the package "center"                      *
* from http://fmwww.bc.edu/RePEc/bocode/cp                *
*****
* The example presented below uses infant birth weight data (Abrevaya 2006) to *
* estimate the interaction between smoking behaviour and maternal age on      *
* birthweight. However, any other data & variables can be defined in the first *
* section of the do-file. The code in its current form relies on continuously *
* scaled or dummy-variables as moderators                                   *
*****

set more off
version 13
*****Data and Variables*****

*path to data
global data "http://www.stata-press.com/data/mlmus2/smoking"

*define unitvar
global i "momid"

*define timevar
global t "idx"

*define time-varyig interactors
local x "mage"
local z "smoke"

*define dependent variable
local y "birwt"

*define covariates
local w "married"

use $data, clear
xtset $i $t
des
xtdes
preserve
center `y' `x' `z' `w', inplace // grand mean centering enables comparison of main effects across models

*****Estimate Classic FE Interaction*****
```

```

xtreg `y' c.`x'##c.`z' `w', fe
*or via
generate int_`x'`z'=`x'*`z'
tempvar sample
xtreg `y' `x' `z' int_`x'`z' `w', fe
estimates store FE_IE
generate `sample'=1 if e(sample)==1

*****Estimate double demeaned (within-unit) interaction*****
foreach var of varlist `x' `z'{
egen mean`var'=mean(`var') if `sample'==1, by ($i)
generate dm`var'=`var'-mean`var'
}
replace int_`x'`z' = dm`x' * dm`z'
qui xtreg `y' `x' `z' int_`x'`z' `w', fe
estimates store dd_IE
di _n in gr "Fixed-effects regression with double-demeaned interaction xz"
est replay dd_IE

* Compare FE_IE,dd_IE
est tab FE_IE dd_IE , b(%9.5f) se(%9.5f) p(%9.8f)

*****Hausman-test on systematic differences between models with standard fe-****
*****and "real" within-estimator of interaction.*****
noisily hausman dd_IE FE_IE
restore

```

A5. Formal Prove that Double-Demeaned Estimator does not Work with t = 2

For each pair of values, the absolute difference between the values is twice as high as the absolute difference between each of the values and the mean of the two values: if $t=2$, for each i and each variable a it is

$$(A4.1) \quad 2 * (\bar{a}_i - a_{it})^2 = (a_{i1} - a_{i2})^2.$$

Therefore,

$$(A4.2) \quad var(a_{i1}; a_{i2}) = \left(\frac{a_{i1} - a_{i2}}{2}\right)^2.$$

Now let a be the product of two demeaned variables and thus $a_{i1} - a_{i2} = dm(x_{i1})dm(z_{i1}) - dm(x_{i2})dm(z_{i2})$. Then $var(a_{i1}; a_{i2}) = 0$, as $dm(x_{i1})dm(z_{i1}) - dm(x_{i2})dm(z_{i2})$ can be written as

$$(A4.3) \quad \left(x_{i1} - \frac{x_{i1} + x_{i2}}{2}\right) * \left(z_{i1} - \frac{z_{i1} + z_{i2}}{2}\right) - \left(x_{i2} - \frac{x_{i1} + x_{i2}}{2}\right) * \left(z_{i2} - \frac{z_{i1} + z_{i2}}{2}\right)$$

$$(A4.4) \quad = \frac{-(x_{i1}z_{i2} + x_{i2}z_{i1})}{2} + \frac{(x_{i1} + x_{i2})(z_{i1} + z_{i2})}{4} - \frac{-(x_{i2}z_{i1} + x_{i1}z_{i2})}{2} - \frac{(x_{i1} + x_{i2})(z_{i1} + z_{i2})}{4} = 0$$

Thus, for $t=2$, realizations of demeaned products are constant within each i . Accordingly, double-demeaned products are fixed at zero. Because of (A5.1), for $t=2$ it also is $dm(y_{i1}) = -dm(y_{i2})$. Consequently, the unit specific slopes of $dm(y)$ on $dm(x)*dm(z)$ are zero. Therefore, even for an undemeaned product of demeaned variables a coefficient can't be identified if $t = 2$.

A6. Replications (Full Results)

Tables A6(1) to A6(6) outline the full results of the replicated models. Thus, this supplement complements the concentrated presentation of replicated models in Table 1.

Variable names have been adopted from the original studies. Generally, we report the models as they were outlined and formatted by the authors. We refrain, however, from reporting divergent schemes of indicating significance and therefore just report coefficients and standard errors. Some authors included additional controls which are not reported in their main document, and therefore do also not appear in our tables.

In each table, the first column of coefficients cites FE results from the original studies. The second column outlines our replication of the standard FE model. Some reproduced results differ from the published results:

- Oesch and Lipps (2013) did not use the standard FE approach, but specified a random effects model with demeaned time-varying independent variables. As the authors did not demean "quasi" time-constant variables with residual intra-individual variation (like education and region), covariance patterns among the independent variables differ from standard FE. This explains the differences between columns 1 and 2 in Tables A6(2) and A6(3).
- For the replication of the analyses of Kühirt (2012), we used a different (newer) version of the SOEP than the author. Hence, original figures (column 1) and replicated original analyses (column2) in Table A6(4) slightly differ.
- For the replication of Killewald and Gough (2013), we used the same robust estimation of standard errors as the authors.
- The models in Column 3 are equivalent to those reported in Column 2, but are based on grand mean centered variables. As a consequence, some coefficients of main effects seem to differ, but are actually just measured at different reference values of interacted variables (concretely, the main coefficients in Column 3 refer to the mean, not the zero point of interacted variables).

Column 4 reports the coefficients of FE-Models with double-demeaned interactions (dd-IE).

Table A6(1). Replication of Schunck (2013): The interaction between mother's age (mage) and smoking (smoke) on birth weight

	Standard FE (Original)		Standard FE (Replication)		Standard FE (Mean Centered)		Double Demeaning	
	B	SE	B	SE	B	SE	B	SE
<i>Main Effects</i>								
smoke	-205.66	134.01	-205.66	134.01	-97.17	31.57	-102.59	29.58
mage	22.75	3.09	22.75	3.09	23.28	3.06	23.12	3.05
<i>Interaction</i>								
smoke*mage	3.79	4.96	3.79	4.96	3.79	4.96	-78.98	46.81
Total observations (n)	8604		8604		8604		8604	
N (with t>1)	3978		3978		3978		3978	
N (with t>2)	648		648		648		648	

Based on NCHS-Data 1990-1998 (Abrevaja 2006)

The original Model is outlined in Schunck (2013), Table 2(1)

Table A6(2). Replication of Osch/Lipps (2013): The interaction between individual and ambient unemployment on life satisfaction (Women in Germany)

	Standard FE (Original)		Standard FE (Replication)		Standard FE (Mean Centered)		Double Demeaning	
	B	SE	B	SE	B	SE	B	SE
<i>Main Effects</i>								
Age	-0.077	0.006	-0.066	.004	-0.066	.004	-0.066	.004
Age squared	0.0004	0.00007	0.0004		0.0004	.00004	0.0004	.00004
Below upper secondary education (ref.)								
Upper secondary education	0.142	0.027	0.066	0.026	0.066	0.026	0.065	0.026
Tertiary education	0.335	0.033	0.204	0.034	0.204	0.034	0.203	0.034
Lives with a partner	0.123	0.026	0.121	0.016	0.121	0.016	0.121	0.017
Annual household income	0.223	0.012	0.231	0.009	0.231	0.009	0.229	0.009
Employed and not unemployed year after (ref.)								
Employed year before becoming unemployed	-0.196	0.028	-0.199	0.024	-0.199	0.024	-0.194	0.023
1st yearUnemployed	-0.403	0.124	-0.406	0.089	-0.557	0.084	-0.545	0.089
2nd yearUnemployed	-0.319	0.131	-0.319	0.093	-0.470	0.088	-0.474	0.085
Spell of unemployment	-0.003	0.063	0.001	0.045	0.001	0.045	-0.015	0.045
Spell of unemployment squared	0.000	0.007	0.000	0.053	0.000	0.053	0.001	0.005
Regional unemployment rate	-0.028	0.003	-0.034	0.003	-0.035	0.003	-0.036	0.003
<i>Interaction</i>								
Unemployed x regional unemployment rate	-0.017	0.005	-0.017	0.003	-0.017	0.003	-0.006	0.010
Total observations (n)	126925		126925		126925		126925	
N (with t>1)	11730		11730		11730		11730	
N (with t>2)	10361		10361		10361		10361	

Based on SOEP-Data (1984-2010)

The original FE model is outlined in Osch and Lipps (2013), Table 1(3)

Table A6(3). Replication of Osch/Lipps (2013): The interaction between individual and ambient unemployment on life satisfaction (Men in Germany)

	Standard FE (Original)		Standard FE (Replication)		Standard FE (Mean Centered)		Double Demeaning	
	B	SE	B	SE	B	SE	B	SE
<i>Main Effects</i>								
Age	-0.071	0.006	-0.066	0.004	-0.066	0.004	-0.066	0.004
Age squared	0.00031	0.00006	0.0003	0.00004	0.0003	0.00004	0.0003	0.00004
Below upper secondary education (ref.)								
Upper secondary education	0.082	0.025	0.002	0.025	0.002	0.025	0.002	0.025
Tertiary education	0.318	0.031	0.137	0.034	0.137	0.034	0.136	0.034
Lives with a partner	0.208	0.024	0.208	0.016	0.208	0.016	0.208	0.016
Annual household income	0.208	0.012	0.220	0.009	0.220	0.009	0.219	0.009
Employed and not unemployed year after (ref.)								
Employed year before becoming unemployed	-0.252	0.028	-0.253	0.022	-0.253	0.022	-0.255	0.022
1st yearUnemployed	-0.715	0.122	-0.719	0.082	-0.663	0.077	-0.659	0.077
2nd yearUnemployed	-0.585	0.127	-0.586	0.084	-0.531	0.079	-0.525	0.079
Spell of unemployment	-0.172	0.065	-0.169	0.041	-0.169	0.041	-0.161	0.041
Spell of unemployment squared	0.021	0.008	0.021	0.005	0.021	0.005	0.020	0.005
Regional unemployment rate	-0.026	0.003	-0.036	0.003	-0.036	0.003	-0.036	0.003
<i>Interaction</i>								
Unemployed x regional unemployment rate	0.006	0.005	0.006	0.003	0.006	0.003	-0.030	0.010
Total observations (n)	144261		144261		144261		144261	
N (with $t > 1$)	14148		14148		14148		14148	
N (with $t > 2$)	12426		12426		12426		12426	

Based on SOEP-Data (1984-2010)

The original FE model is outlined in Osch and Lipps (2013), Table 1(4)

Table A6(4). Replication of K uhirt (2012): The interaction between number of children and logged income on time spent in childcare (Men in Germany).

	Standard FE (Original)		Standard FE (Replication)		Standard FE (Mean Centered, Regular SEs)		Double Demeaning	
	B	SE	B	SE	B	SE	B	SE
<i>Main Effects</i>								
No children (ref.)								
One child	2.01	0.09	1.99	0.09	1.99	0.05	2.00	0.06
Two or more children	1.98	0.11	1.93	0.11	1.93	0.07	1.91	0.07
Age of youngest child	0.07	0.03	0.11	0.03	0.11	0.02	0.10	0.02
Age of youngest child ²	-0.01	0.00	-0.01	0.00	-0.01	0.00	-0.01	0.00
One child * Age of youngest child	-0.05	0.04	-0.09	0.04	-0.09	0.03	-0.09	0.03
One child * Age of youngest child ²	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.00
Ln(yearly HH-income)	0.07	0.06	0.06	0.07	-0.14	0.04	-0.11	0.05
<i>Interaction</i>								
One child * Ln(yearly HH-income)	-0.29	0.13	-0.32	0.13	-0.32	0.08	0.07	0.13
Two or more children * Ln(yearly HH-income)	-0.51	0.15	-0.43	0.16	-0.43	0.09	-0.02	0.13
Total observations (n)	13921		12872		12872		12872	
N (with t>1)			1527		1527		1527	
N (with t>2)			1331		1331		1331	

Replications Based on SOEP-Data v31 (1984-2007)

The original FE Model is outlined in K uhirt 2012, Table A2(4).

Table A6(5). Replication of Killwald and Gough (2013): The interaction between marital status and having children on logged wage (Men in the USA)

	Standard FE (Original)		Standard FE (Replication)		Standard FE (Mean Centered)		Double Demeaning	
	B	SE	B	SE	B	SE	B	SE
<i>Main Effects</i>								
Single (ref.)								
Married	0.073	0.010	0.073	0.010	0.098	0.010	0.077	0.010
Cohabiting	0.058	0.011	0.058	0.011	0.058	0.011	0.044	0.011
Divorced	0.020	0.014	0.020	0.014	0.020	0.014	-0.009	0.014
Childless (ref.)								
1 Child	0.010	0.014	0.010	0.014	0.014	0.010	0.018	0.009
2+ Children	-0.034	0.018	-0.034	0.018	0.003	0.013	0.032	0.012
<i>Interaction</i>								
1 Child * Married	0.009	0.015	0.009	0.015	0.009	0.015	-0.041	0.021
2+ Children * Married	0.082	0.017	0.082	0.017	0.082	0.017	-0.032	0.022
Total observations (n)	56403		56403		56403		56403	
N (with $\tau > 1$)	4411		4411		4411		4411	
N (with $\tau > 2$)	4308		4308		4308		4308	

Based on NLSY-Data (1979-2008)

The original FE model is outlined in Killwald and Gough (2013), Table 1(4)

Table A6(6). Replication of Killwald and Gough (2013): The interaction between marital status and having children on logged wage (Women in the USA)

	Standard FE (Original)		Standard FE (Replication)		Standard FE (Mean Centered)		Double Demeaning	
	B	SE	B	SE	B	SE	B	SE
<i>Main Effects</i>								
Single (ref.)								
Married	0.037	0.011	0.037	0.011	0.030	0.012	0.026	0.012
Cohabiting	0.036	0.012	0.036	0.012	0.036	0.036	0.034	0.012
Divorced	0.045	0.016	0.045	0.016	0.045	0.045	0.041	0.016
Childless (ref.)								
1 Child	-0.051	0.017	-0.051	0.017	-0.057	0.011	-0.059	0.011
2+ Children	-0.137	0.023	-0.137	0.023	-0.145	0.017	-0.147	0.015
<i>Interaction</i>								
1 Child * Married	-0.012	0.018	-0.012	0.018	-0.012	0.018	-0.031	0.025
2+ Children * Married	-0.016	0.021	-0.016	0.021	-0.016	0.021	-0.057	0.028
<i>Constant</i>			2.25		0		0	
Total observations (n)	46240		46240		46240		46240	
N (with $t > 1$)	3915		3915		3915		3915	
N (with $t > 2$)	3805		3805		3805		3805	

Based on NLSY-Data (1979-2008)

The original FE model is outlined in Killiwald and Gough (2013), Table 1(4)